EGIInet Supplementary Materials

Hang Xu[1*], Chen Long[1*], Wenxiao Zhang[2†], Yuan Liu[3], Zhen Cao[1], Zhen Dong[1], and Bisheng Yang[1]

[1] LIESMARS, Wuhan University
{190107xh, chenlong107, zhen.cao, dongzhenwhu,bshyang}@whu.edu.cn
[2] University of Science and Technology of China
wenxxiao.zhang@gmail.com
[3] The University of Hong Kong
yuanly@connect.hku.hk

A Implementation Details

Input partial point cloud contains 2048 points, and input single view image is a $224 \times 224 \times 3$ RGB image. The output complete point cloud contains 2048 points and is evaluated with ground truth point cloud of 2048 points.

The train-test split follows the list provided by ViPC [6]. For each object, ShapeNet-ViPC [6] provides 24 different views corresponding to different missing situations. During training and testing, pairs of point clouds and images are randomly fed to the model. This means that a partial point cloud may be assisted by different images in different training epochs.

We use the Adam optimizer with an initial learning rate of 0.001 to train our model. We train our model for 160 epochs. The learning rate decreased by 70% every 16 epochs.

During tokenization, the point cloud is downsampled to 256 points and image is divided by a 16×16 grid. Therefore the number of point cloud tokens and image tokens are both 256. Each token in the token feature has 192 channels. The token features keep the size of 256×192 when passing the SFE and SFTnet. The fused feature is also the size of 256×192, and we use this fused feature to infer the missing part of the point cloud.

B Results on Real Scenes

We also report qualitative results on KITTI [2] cars extracted by [5]. Qualitative results are shown in Fig. 1. Though there is a synthetic-to-real gap, our method can still able to give a reasonable prediction.

C Comparisons about Model Sizes

We compare the size of our model with existing view-guided completion models [1, 6, 7] in table 1. The comparison results show that our model achieves the best results (+ **16% CD**) with the smallest number of parameters.

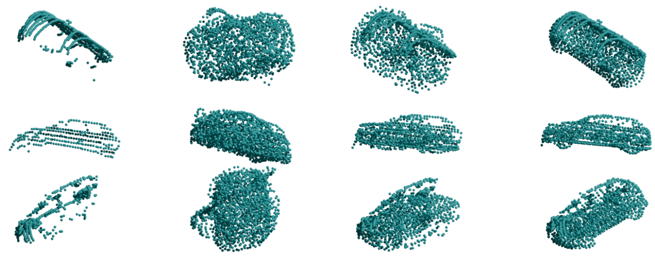

Fig. 1: Qualitative results on KITTI [2] cars.

Methods	Number of parameters(M)	$CD \times 10^3$
ViPC [6]	17.43	3.308
CSDN [7]	16.85	2.570
XMFnet [1]	9.57	1.443
Ours	**9.03**	**1.211**

Table 1: Comparisons on model sizes and performance.

D Comparison with Single View Reconstruction Method

Compared with the emergence of ViPC [6] in 2021, the current single view reconstruction has made a great breakthrough. To dispel the doubts about why we still need to study view guided point cloud completion instead of directly reconstructing shapes from images, we made a qualitative comparison with the latest single-view reconstruction model TripoSR [4] as fairness as possible. TripoSR is a 3D reconstruction model based on LRM (Large Reconstruction Model) [3] that can produce a 3D mesh from a single view image. Due to the need for a strong prior and smooth distribution of latent space for robust single view reconstruction, it can only restore a roughly reasonable shape but difficult to achieve fine surface reconstruction. In addition, single view reconstruction method relies heavily on the quality of the image. On the other hand, our method can achieve fine reconstruction of the missing parts based on the point cloud prior and supplemented by finely extracted image structure information. we uniformly sample 2048 points on the mesh surface and normalized the sampled point cloud into a unit sphere. Qualitative comparisons are shown in Fig. 2. It can be seen that TripoSR [4] tends to generate a shape with the outline of the image but lacks accurate depth on the image in the ShapeNet-ViPC dataset [6] in some categories. This indicates that it is unreliable to predict the complete shape based on the image alone. The point cloud completion task can use the image to infer the complete shape more reliably on the basis of the partial point cloud. Further more, relying on images to infer possible shapes is much more costly than point cloud completion. View guided point cloud completion [6] is a relatively cost-effective and reliable option.

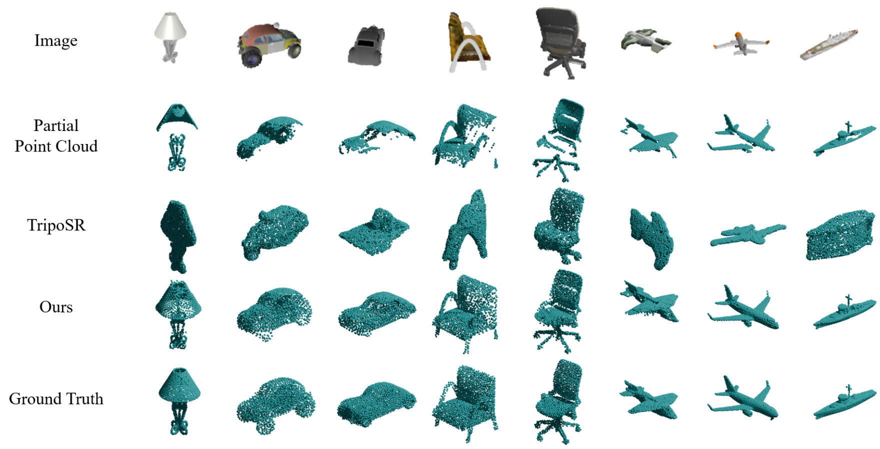

Fig. 2: Qualitative comparison with TripoSR [4]. It is difficult for single-view reconstruction to accurately estimate depth and predict the back in poor-quality images

E Additional Qualitative Results

We provide additional completion results of our model in Fig. 3, Fig. 4, Fig. 5 and Fig. 6. All the point clouds in the figures are artificially rotated to an axis-aligned angle for easy observation. It can be seen that our method can give reasonable prediction in a variety of missing cases. Our model perform well on categories with large individual differences, such as lamps, chairs, etc. This is due to a deeper understanding of structural information, rather than learning the average shape distribution of the dataset.

Fig. 3: Additional results on watercraft and table.

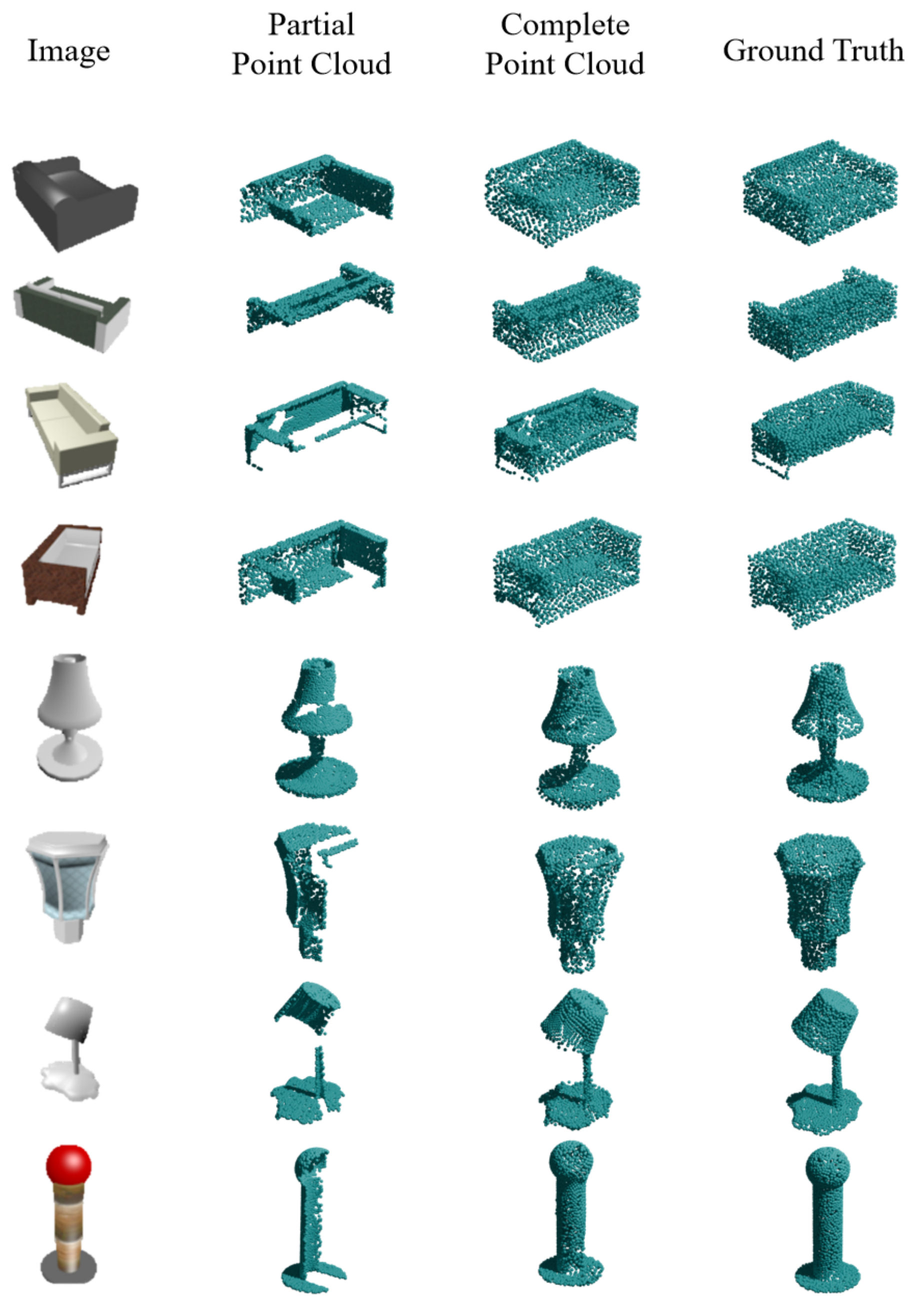

Fig. 4: Additional results on sofa and lamp.

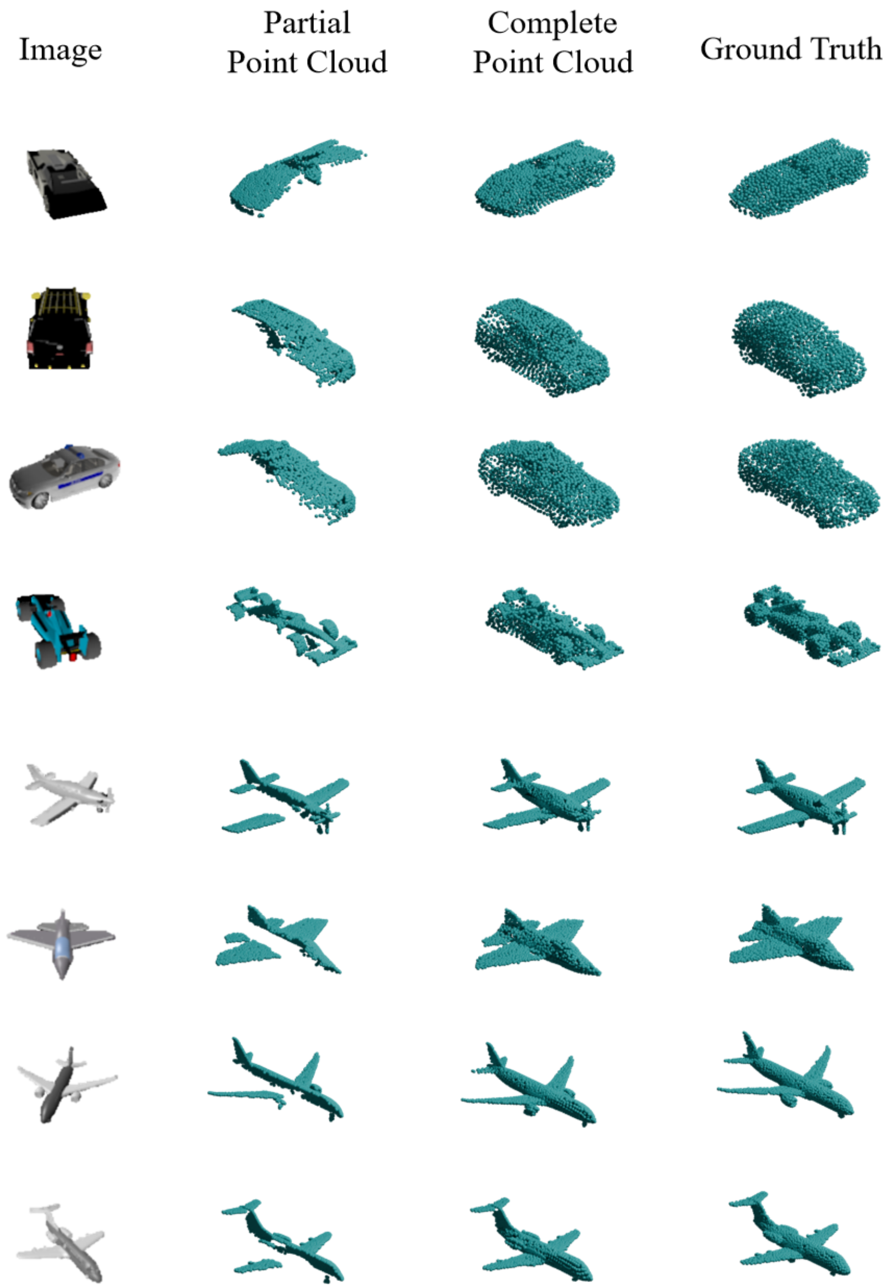

Fig. 5: Additional results on car and plane.

Fig. 6: Additional results on chair and cabinet.

References

1. Aiello, E., Valsesia, D., Magli, E.: Cross-modal learning for image-guided point cloud shape completion. Advances in Neural Information Processing Systems **35**, 37349–37362 (2022)
2. Geiger, A., Lenz, P., Stiller, C., Urtasun, R.: Vision meets robotics: The kitti dataset. The International Journal of Robotics Research **32**(11), 1231–1237 (2013)
3. Hong, Y., Zhang, K., Gu, J., Bi, S., Zhou, Y., Liu, D., Liu, F., Sunkavalli, K., Bui, T., Tan, H.: Lrm: Large reconstruction model for single image to 3d. arXiv preprint arXiv:2311.04400 (2023)
4. Tochilkin, D., Pankratz, D., Liu, Z., Huang, Z., , Letts, A., Li, Y., Liang, D., Laforte, C., Jampani, V., Cao, Y.P.: Triposr: Fast 3d object reconstruction from a single image. arXiv preprint arXiv:2403.02151 (2024)
5. Yuan, W., Khot, T., Held, D., Mertz, C., Hebert, M.: Pcn: Point completion network. In: 2018 international conference on 3D vision (3DV). pp. 728–737. IEEE (2018)
6. Zhang, X., Feng, Y., Li, S., Zou, C., Wan, H., Zhao, X., Guo, Y., Gao, Y.: View-guided point cloud completion. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 15890–15899 (2021)
7. Zhu, Z., Nan, L., Xie, H., Chen, H., Wang, J., Wei, M., Qin, J.: Csdn: Cross-modal shape-transfer dual-refinement network for point cloud completion. IEEE Transactions on Visualization and Computer Graphics (2023)